DOCUMENT RESUME

ED 460 162 TM 033 637

AUTHOR Pascoe, Donna; Halpin, Glennelle
TITLE Legal Issues To Be Considered When Testing Teachers for Initial Licensing.
PUB DATE 2001-11-00
NOTE 18p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (30th, Little Rock, AR, November 14-16, 2001).
PUB TYPE Information Analyses (070) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Court Litigation; Elementary Secondary Education; Job Skills; *Legal Problems; *Licensing Examinations (Professions); *Preservice Teachers; Reliability; *Teacher Certification; Teacher Qualifications; Test Use; Validity

ABSTRACT

This review covers the test components of validity, reliability, job-relatedness, and test bias in relation to teacher licensing examinations and the legal decisions that have affected policy in this area. The literature provides a history of court decisions and legal rulings that have shaped policy, test design, and test use. The important criteria for implementing a fair and legal program of testing teachers for initial licensure are discussed, along with the guidelines that provide a framework for determining the proper use of tests and other selection procedures. Although the testing of teachers for initial licensure has been challenged on legal grounds purporting discrimination, the courts have upheld the practice of professional employment testing when tests are: (1) not used to discriminate; (2) shown to be job related; and (3) indicative of job qualifications and performance. Criterion-related tests that determine a teacher candidate's competency have become respectable tools for educational improvement. It is not easy to be well-informed about current legal practice with regard to professional licensure, but it is important. (Contains 26 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 460 162

Legal Issues to Be Considered When Testing Teachers for Initial Licensing

Donna Pascoe

Columbus State University

Glennelle Halpin

Auburn University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Pascoe

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

TM033637

Paper presented at the meeting of the Mid-South Educational Research Association,

Little Rock, Arkansas, November 2001

BEST COPY AVAILABLE

Legal Issues to Be Considered When Testing Teachers for Initial Licensing

Do teachers have the basic minimum competencies to be effective educators? To answer this question, the school reform movement experiencing rapid public acceptance is teacher testing. Standardized criterion-referenced tests provide a quick, easy, and observable measure of whether or not a teacher candidate possesses the basic minimum competencies required to receive initial teacher licensure.

Implementing a testing program is not as simple as one might think. As with all tests, teacher licensing exams must be valid, reliable, fair, and legally defensible. Court decisions have impacted teacher testing by attempting to resolve both technical and social problems associated with testing. Legal actions related to testing have helped to define the direction of the competency-based testing movement. Individuals responsible for teacher-testing programs must have a working understanding of what makes a test valid, reliable, and unbiased as well as an understanding of how and when testing may be used.

This review covers the test components of validity, reliability, job-relatedness, and test bias as determined by impact for teacher licensing exams. The literature provides a history of court decisions and legal rulings that have shaped policy, test design, and use. A review of the important criteria for implementing a fair and legal testing program will be discussed, along with guidelines that provide a framework for determining the proper use of tests and other selection procedures. Licensing and certification boards from State Departments of Education would benefit from this review of literature as it pertains to testing requirements determined by federal law and psychometric practices.

Definitions

The terms licensure and certification have been used interchangeably within state organizations, creating much confusion. The Carnegie Task Force on Teaching as a Profession determined that a distinction should be made between the terms licensure and certification. The task force defined licensure as a function of the state acting on its authority to protect and promote the general welfare, and it defined certification as a function of the profession acting to acknowledge those who demonstrate advanced capabilities (Carnegie Forum, 1986). Put more simply, licensure is carried out by the state agency and certification is awarded by an external professional body. The formation of the National Board for Professional Teaching Standards (NBPTS) in 1987 emphasized the distinction between the two. For this review the following definitions will be used.

Licensing: Teacher licensing is a process that guarantees to the public that individuals have met the established requirement. It is intended that those receiving a license to teach possess a minimum level of knowledge and skills needed to function in the role designated by the state license (Darling-Hammond, Wise, & Klein, 1999; Green, 1986; Roth, 1996; Sykes, 1990).

Certification: Teacher certification is a process that goes beyond minimum competency using more advanced standards established by the profession. These are often performance based (Baratz-Snowden, 1991; Darling-Hammond, 1998; Kowalski, 1988). Certification is controlled by occupational groups that develop and maintain advanced competence in the field of practice. Further, certification is voluntary and does not involve restrictions on the right to practice (Roth, 1996; Sykes, 1990). Those becoming certified are experienced teachers.

Licensing of Teachers

In the case of teacher testing for initial licensing, the intent for all states is to determine whether or not a candidate has the minimum basic competencies to perform in the capacity of teacher. The Standards for Educational and Psychological Testing (AERA/APA/NCME, 1999) state the following:

> The primary purpose of licensure or certification is to protect the public. Licensing requirements are imposed to ensure that those licensed posses knowledge and skills in sufficient degree to perform important occupational activities safely and effectively. (p.156)

The specific knowledge and skills needed by the minimally competent professional are determined by those possessing expertise in the domains being tested. A combination of professional judgement and empirical evidence is used to decide what a minimally competent beginning teacher would need in order to perform safely and effectively on the job. These criteria for success will be different according to the professionals making recommendations to each State Department of Education issuing a license.

Each reform movement inspires newer, different, and more precise testing for accountability. Once a State Department of Education makes the decision to test its teachers, the determination must be made as to what test is a valid measure of teaching ability and which method of standard setting is to be used. McDonough and Wolf (1988) identified five issues around which litigation against educational testing programs occurs. These five issues should be considered when states begin to institute teacher-testing programs for initial licensing.

1. The arbitrary and capricious development or implementation of a test or employee selection procedure.
2. The statistical and conceptual validity of a test or procedure.
3. The adverse or disproportionate impact of a testing program or selection procedure on a "protected group."
4. The relevancy of a test or procedure to the identified requirements of the job (job-relatedness).
5. The use of tests or selection procedures to violate an individual's or group's civil rights. (p. 38)

Reform by Testing: Legal Issues

The mere diversity of test types could result in little consistency of use between states. Each state must search for teacher tests that are valid, reliable, fair, and legally defensible. During the 1980s, psychometricians responded to court decisions that impacted teacher testing by attempting to resolve both technical and social problems associated with testing. Legal actions related to testing have helped to define the direction of the competency-based testing movement. No matter what form of assessment a state decides to use, it must be fair and follow legal guidelines. Compliance to fairness of testing practice is the responsibility of many and will depend on informed decision making (AERA et al., 1999).

The two most influential resources and those that provide operational direction for test construction, use, and evaluation of test results are the 1978 revised edition of the Uniform Guidelines on Employee Selection Procedures (1978/2000) and the revised edition of the Standards for Educational and Psychological Testing (AERA et al., 1999). The Uniform Guidelines on Employee Selection Procedures codifies the position of four government agencies

that monitor the use of tests in personnel evaluation: the U.S. Civil Service Commission, the Department of Justice, the Equal Employment Opportunity Commission, and the Department of Labor.

The Guidelines provide a framework for determining the proper use of tests and other selection procedures. Employers, labor organizations, employment agencies, and licensing and certification boards should use the Guidelines in order to comply with requirements of federal law prohibiting employment practices which discriminate on grounds of race, color, religion, sex, and national origin. The Guidelines are not only helpful as a source of reference, but also they are a determinant for legal decisions in cases of litigation. The Guidelines have "considerable force of law" (Thornton, 1998, p. 162) and are a necessary resource for all states making legal determinations to test candidates seeking initial licensure.

The Standards for Educational and Psychological Testing were prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education, providing generally accepted professional standards for evaluating standardized tests and other selection procedures. Standards for Educational and Psychological Testing are not a checklist to evaluate literal compliance but is designed to "promote the sound and ethical use of tests and to provide a basis for evaluating the quality of testing practices" (AERA et al., 1999, p. 1). The Standards provide a frame of reference by which those involved in assessment may evaluate issues pertaining to testing and make judgements in the best interest of all parties. Foremost for both the Guidelines and Standards are the issues of validity, reliability, job-relatedness, and test bias as determined by impact.

Validity and Job-Relatedness

"Validity refers to the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests" (AERA et al., 1999, p. 9). Those using testing as a requirement for initial licensure must gather evidence providing a scientific basis for the intended interpretation of test scores.

The Uniform Guidelines on Employee Selection Procedures refer to three criteria that may be used for validity studies. These are criterion-related validity, construct validity, and content validity. Criterion-related validity would determine selection according to a statistical relationship between scores on a test and a measure of job performance. Criterion validity is concerned with prediction and is not well suited for competency-based teacher certification tests. The courts have suggested that this type of validity study should not be a requirement in teacher tests (Wolf & McDonough, 1987). Because the definition of what attributes make a good teacher is debatable, construct validity (as determined by the courts) has limited application for initial teacher licensure tests (Wolf & McDonough, 1987).

Content validity is most often used in validation studies for teacher licensing tests. The objective of a content validation study has been to determine whether the content provides a representative sample of the domain identified as being necessary to perform the job and is often referred to as job relatedness. The Standards for Educational and Psychological Testing reflect the latest theories and thinking in psychometric measurement. No longer is the tripartite description of validity emphasized (Thornton, 1998). The newer thinking looks at validity as a unitary concept that encompasses an evaluation of information about the adequacy of a test for its stated purpose

and supportable interpretations of test scores. This latest thinking regarding the interpretation of test scores is voiced by Pedhazur and Schmelkin (1991) who stated, "Content validity is not a type of validity at all" (p.79). Content of an instrument, although highly important, is not evidence of the validity of an instrument because validity refers to inferences made about scores, not to an assessment of the content of an instrument (Messick, 1995).

Although there is disparity between the interpretation of content validity between psychometricians and the courts, content validity is the primary evidence used by courts when making decisions about test fairness (Kuehn, Stallings & Holland, 1989; Mehrens & Popham, 1992; Sireci & Green, 2000). Three prominent court cases involving teacher testing resulted in different rulings as determined by validation studies: The United States v. State of North Carolina (1975), United States v. State of South Carolina (1977), and Richardson v. Lamar County Board of Education (1991).

The United States Department of Justice (1975) brought suit against the State of North Carolina for the violation of teacher candidates' rights under the United States Constitution and federal law prohibiting discrimination in employment as described in Title VII of the Civil Rights Act of 1964. The Department of Justice alleged that the state's use of the National Teacher Examinations developed by Educational Testing Service for teacher certification was unfair and discriminatory. The courts ruled that, because a validation study had not been conducted, the use of the examination without validation of the passing score was arbitrary and denied equal protection of law. The court prohibited the state from using the exam for certification (U.S. v. North Carolina, 1975).

The United States Department of Justice (1977) brought suit against the State of South Carolina with identical allegations as in the North Carolina litigation. Unlike the previous case, the United States District Court ruled in South Carolina's favor based on the fact that a validation study of the National Teacher Exam for certification and its passing score had been conducted prior to initiating a testing program (U.S. v. South Carolina, 1975).

The third case decided by the United States District Court was Richardson v. Lamar County Board of Education (1991). Richardson was a teacher who had been terminated by the Board of Education because of her inability to secure a permanent teaching certificate. She had been denied a certificate because of her failure to pass a teacher certification test developed by the National Evaluation Systems (NES) and required by the State of Alabama. The court ruled in favor of the plaintiff determining the State of Alabama's use of the NES exam violated Richardson's rights under Title VII. The Court ruled that the development and validation of the certification test and the process that led to the adoption of the passing score did not meet acceptable and reasonable professional standards. The court ordered that Richardson be reemployed, as well as compensated for back pay and other benefits (Richardson v. Lamar County Board of Education, 1991). This decision by the District Court, once again, was contingent upon validation of the test and its passing scores.

The United States v. State of North Carolina (1975), United States v. State of South Carolina (1977), and Richardson v. Lamar County Board of Education (1991) are three cases that have set a legal precedence regarding the need for validation of tests and their passing scores that must be considered when State Departments of Education begin to use testing programs for teacher certification or licensing.

The Uniform Guidelines on Employee Selection Procedures Sec. 1607, outline the general standards for validity studies, which include the validation and setting of cut scores. The Standards for Educational and Psychological Testing recognize components of professional judgment, as well as the technical and empirical nature of cut scores. With the emphasis on test scores and their interpretation, the techniques for establishing cut scores on tests for initial licensing and certification are of greater consequence and will come under more scrutiny. The establishment of cut scores becomes the standard by which examinees are determined to have passed, receiving a license, or failed, not receiving a license. The high-stakes outcome deems that test content must be demonstrated to be valid and scores set to be reasonable and consistent with expectations of acceptable proficiency for the candidate seeking a teaching license. The Standards recommend that procedures for test selection, creation, validation, and cut score studies be well documented.

Reliability

Reliability is an indication of the degree of consistency in a test's application. The more consistently a test measures the domain it is designed to measure, the more reliable it is. When evaluating teacher tests for certification and licensure, results must be consistent across test items and alternate forms of the test, as well as the decisions made about individuals according to test results (McDonough & Wolf, 1988).

The Uniform Guidelines on Employee Selection Procedures make clear that reliability must be of concern, and that "whenever feasible, appropriate statistical estimates should be made of the reliability of the selection procedure" (p. 212). Biddle (1993) devised a 15-step method for setting cut scores. He suggested statistical procedures recommending that it would be prudent to

calculate the reliability and standard deviation of the licensing test after it is administered using an up-to-date item analysis program that calculates the Horst Modification of the KR-20. The Horst KR-20 corrects for assumption of equal item difficulty (Biddle, 1993). This procedure should be performed by testing contractors, thus providing important information about an exam.

The Standards for Educational and Psychological Testing provide flexibility in the manner chosen to report reliability. Reliability information may be reported in terms of "variances or standard deviations of measurement errors, in terms of one or more coefficients, or in terms of IRT-based test information factions" (p. 27). Reliability measures should help insure that an examinee would consistently earn a comparable score on the exam or on other forms of the exam.

Impact and Bias

The Uniform Guidelines on Employee Selection Procedures are designed to protect individuals against adverse impact to employment because of testing programs. Adverse impact is said to occur when any race, sex, or ethnic group has a passing rate of less than 80% of the passing rate for the group with the highest rate (section 4D). This criterion is sometimes referred to as the "80% Rule of Thumb." However, it may still constitute adverse impact if smaller differences occur that are significant in practical or statistical terms. Statistical significance is set between two and three standard deviations and not a minimum probability level (Biddle, 1993). The evidence of bias or lack of bias is sought in the relationships between test and criterion scores. An accepted definition is presented by the Standards for Educational and Psychological Testing as follows: "No bias exists if the regression equations relating the test and the criterion are indistinguishable for the groups in question" (AERA et al., 1999, p. 79).

The most recent test case involving negative impact and bias was Association of Mexican-American Educators v. California (1996). Several advocacy groups and a small group of individuals brought a class action suit against the State of California and the California Commission on Teacher Credentialing, alleging that the use of the California Basic Skills Test (CBST) for credentialing violated their civil rights under Titles VI and VII. In making rulings about the test content and the passing score, it was determined that (a) the CBST was job-related and consistent with the Commission's teacher certification responsibilities; (b) the passing scores reflected the reading, writing, and mathematics skills appropriate for entry-level teachers; and (c) the plaintiffs had failed to prove the existence of valid reading, writing, and math skill measurement alternatives having less adverse impact on racial and ethnic minorities (Association of Mexican-American Educators v. California, 1996).

Records are to be kept of testing results as they affect Blacks, American Indians, Alaskan Natives, Asians, Pacific Islanders, Hispanics (including Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish origin or culture regardless of race), and Whites (Uniform Guidelines, 1978). If adverse impact is detected, the cause must be investigated to ascertain whether bias is present in the test items, content and job-relatedness, administration, or passing score. These records should be part of validation and setting of passing scores for all teacher certification programs. The Civil Rights Act of 1991 specifically requires that those instituting testing requirements prove that it is a necessary standard for successful job performance and that it be job related.

In all litigation research, the State of Alabama plays a prominent role in the historical evolution of the development of non-discriminate teacher testing. In response to political

pressures to reduce adverse impact of teacher exams, Alabama instituted the Golden Rule formula and extended it to rule out the use of all test items having substantial adverse impact on minority candidates. The Educational Testing Service feared that the use of the strict Alabama version of the Golden Rule would become the norm for all other states, thus rendering it a threat by compromising the basic integrity of teacher certification testing (Phillips, 1991; Rebell, 1988). A test must be designed to separate those who have the minimum basic skills from those who do not possess the basic skills at the time of testing. Currently the State of Alabama is reinstituting teacher testing after many years.

Conclusion

Although the testing of teachers for initial licensure has been challenged on legal grounds purporting discrimination, the courts have upheld the practice of professional employment testing when tests (a) are not used to discriminate, (b) are shown to job related, and (c) indicatives of job qualifications and performance (LaMorte, 2002).

The history of teacher testing indicates that testing regimes are here to stay. State-mandated, high-stakes testing programs continue to represent order and control to those demanding educational accountability. The public wants and deserves well qualified teachers. An initial step toward actualizing this goal is to determine whether or not candidates have basic minimum competencies and the professional knowledge needed to teach. Criterion tests that determine a teacher candidate's competency have become respectable tools for educational improvement. Challenges in the future will surely work to refine and redefine the appropriate and legal uses of teacher testing for initial licensure. It is up to all those concerned to be well informed about current legal practice in the area of competency testing for professional licensure. This

directive is not an easy one as the political nature of testing contributes to the rapidly changing legislation surrounding testing programs.

References

Association of Mexican-American Educators v. California (1996). *Association of Mexican-American Educators v. California*, 937 F. Supp. 1397 (N.C. Cal. 1996), aff'd, 183 F.3d 1055 (9th Cir. 1999).

American Educational Research Association, American Psychological Association, & National Council for Measurement in Education (AERA/APA/NCME), Joint Committee on Educational Psychological Testing, (1999). *Standards for educational and psychological testing* (2nd ed.). Washington, DC: American Educational Research Association.

Baratz-Snowden, J. (1991). Performance assessments for identifying excellent teachers: The National Board for Professional Teaching Standards charts its research and development course. *Journal of Personnel Evaluation in Education*, 5, 133-145.

Biddle, E. R. (1993). How to set cutoff scores for knowledge tests used in promotion, training, certification, and licensing. *Public Personnel Management*, 22, 63-79.

Carnegie Forum on Education and the Economy. (1986). *A nation prepared: Teachers for the 21st Century*. Washington, DC: Author.

Darling-Hammond, L. (1998). Teacher and teaching: Testing policy hypotheses from a national commission report. *Educational Researcher*, 27, 5-15.

Darling-Hammond, L., Wise, A. E., & Klein, S. P. (1999). A license to teach: Raising standards for teaching. San Francisco: Jossey-Bass.

Green, J. (Ed.). (1986 July). *What next? More leverage for teachers*. Denver: Education Commission of the States.

Kowalski, T. J. (1988 Spring). One case for national certification of teachers. *The Teacher Educator,* 23(4), 2-9.

Kuehn, P.A., Stallings, W.M., & Holland. C.L., (1989, March). *Court-defined job analysis standards in content validation,* Paper presented at the National Council on Measurement in Education, San Francisco.

LaMorte, M. W. (2002). *School law: Cases and concepts* (7th ed.) Boston: Allyn & Bacon.

McDonough, W., & Wolf, W.C. (1988). Court actions which helped define the direction of the competency-based testing movement. *Journal of Research and Development in Education,* 21, 37-43.

Mehrens, W. A., & Popham, W. J. (1992). How to evaluate the legal dependability of high-stakes tests. *Applied Measurement in Education,* 5, 265-283.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into scoring meaning. *American Psychologist,* 9, 741-749.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum.

Phillips, S. E. (1991). The golden rule remedy for disparate impact on standardized testing: Progress or regression? *West's Education Law Reporter,* 63, 383-428.

Rebell, A. M. (1988). Legal issues concerning bias in testing. In R. G. Allen, P. M. Nassif, & S. M. Elliot (Eds.), *Bias issues in teacher certification testing* (pp. 1-18). Hillsdale, NJ: Lawrence Erlbaum.

Richardson v. Lamar County Board of Education et al. 729 F. Supp. 806, 820-821 M.D. Ala. (1989), aff'd 935 F. 2d 1240 (5th Circuit 729 F.1991).

Roth, A. R. (1996). Standards for certification, licensure, and accreditation. In J. P. Sikula, T. J. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education: A project of the association of teacher educators* (2nd ed., pp. 242-278). New York: Macmillan.

Sireci, S. G., & Green III, P. C. (2000). Legal and psychometric criteria for evaluation teacher certification tests. *Educational Measurement: Issues and Practice,* 19(1), 22-34.

Sykes, G. (1990). Licensure and certification of teachers: An appraisal. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 62-75). Newbury Park: Sage.

Thornton III, G. C. (1998). Committee revises standards for educational and psychological testing: Implications for employment testing. *Employment Testing: Law & Policy Reporter,*7, 161-165.

Uniform Guidelines on Employee Selection Procedures (1978/2000), Code of Federal Regulations, Title 29, Volume 4, Parts 900 to 1899. Revised as of July 1, 2000. Part 1607 [Online]. Available: http://frwebgate.access.gpo,gov/cgi-bin/ge

U.S. v. North Carolina, 400 F. Supp. 343 (E.D.N.C., 1975)

U.S. v. South Carolina, 445 F. Supp. 1094 (D.S.C., 1977), aff'd, 434 U.S. 1026 (1978).

Wolf, W.C. & McDonough, W. (1987). Testing teachers: Legal and psychometric considerations. *Educational Policy, 1,* 199-214.